

**LOAD LEVELING IN MOBILE AD-HOC NETWORKS TO SUPPORT END-TO-END DELAY REDUCTION, QoS AND ENERGY LEVELING**

**Field of the Invention**

**[0001]** The present invention relates to the field of communications networks, and, more particularly, to mobile ad hoc networks and related methods.

**Background of the Invention**

**[0002]** Wireless networks have experienced increased development in the past decade. One of the most rapidly developing areas is mobile ad-hoc networks (MANETs). Physically, a MANET includes a number of geographically-distributed, potentially mobile nodes sharing one or more common radio channels. Compared with other types of networks, such as cellular networks or satellite networks, the most distinctive feature of MANETS is the lack of any fixed infrastructure. The network is formed of mobile and stationary nodes, and is created on the fly as the nodes communicate with each other. The network does not depend on a particular node and dynamically adjusts as some nodes join or others leave the network.

**[0003]** In a hostile environment where a fixed communication infrastructure is unreliable or

unavailable, such as in a battle field or in a natural disaster area struck by earthquake or hurricane, a MANET can be quickly deployed to provide much needed communications. While the military is still a major driving force behind the development of these networks, ad-hoc networks are quickly finding new applications in civilian or commercial areas. MANETs will allow people and applications to exchange data in the field or in a class room without using any network structure except that which they create by simply turning on their computers or PDAs.

**[0004]** As wireless communication increasingly permeates everyday life, new applications for MANETs will continue to emerge and become an important factor in wireless communications. Yet, MANETs pose serious challenges to designers. Due to the lack of a fixed infrastructure, nodes must self-organize and reconfigure as they move, join or leave the network. All nodes are essentially the same, and there is no natural hierarchy or central controller in the network. All functions have to be distributed among the nodes. Nodes are often powered by batteries and have limited communication and computation capabilities. Also, the bandwidth of the system is usually limited. The distance between two nodes often exceeds the radio transmission range, and a transmission may have to be relayed by other nodes before reaching its destination. Consequently, a MANET network typically has a multi-hop topology, and this topology changes as the nodes move around.

**[0005]** The MANET working group of the Internet Engineering Task Force (IETF) has been actively evaluating and standardizing routing protocols, including

multicasting protocols. Because the network topology changes arbitrarily as the nodes move, information is subject to becoming obsolete, and different nodes often have different views of the network, both in time (information may be outdated at some nodes but current at others) and in space (a node may only know the network topology in its neighborhood and not far away from itself).

**[0006]** A routing protocol needs to adapt to frequent topology changes, possibly with less than accurate information. Because of these unique requirements, routing in these networks is very different than in others. Gathering fresh information about the entire network is often costly and impractical. Some routing protocols are reactive (i.e., on-demand) protocols. That is, they collect routing information only when necessary and only to destinations to which they need routes, and do not maintain unused routes. In this way the routing overhead may be reduced compared to pro-active protocols, which maintain optimal routes to all destinations at all time. Ad Hoc on Demand Distance Vector (AODV), Dynamic Source Routing (DSR) and Temporally Ordered Routing Algorithm (TORA) are representatives of reactive routing protocols presented at the MANET working group.

**[0007]** An example of a proactive routing protocol is Optimal Link State Routing (OLSR). Examples of other various routing protocols include Destination Sequenced Distance-Vector (DSDV) routing which is disclosed in U.S. Patent No. 5,412,654 to Perkins, and the Zone Routing Protocol (ZRP) which is disclosed in U.S. Patent No. 6,304,556 to Haas. ZRP is a hybrid protocol using both proactive and reactive approaches.

**[0008]** These conventional routing protocols use a best effort approach in selecting a route from the source node to the destination node. Typically, minimizing the number of hops is the main criteria in such approaches.

**[0009]** Quality-of-service (QoS) routing in MANETs is gaining interest. To provide quality-of-service, a protocol needs not only to find a route but also to identify and/or secure the resources along the route. Because of the potentially limited, shared bandwidth of the network, and the lack of a central controller which can account for and control these limited resources, nodes must negotiate with each other to manage the resources required for QoS routes. This is further complicated by frequent topology changes. Due to these constraints, QoS routing is more demanding than best-effort or minimum-hop routing.

**[0010]** Some examples of QoS routing approaches are set forth by Chenxi Zhu in the publication entitled "Medium Access Control and Quality-of-Service Routing for Mobile Ad Hoc Networks," 2001, and by M. Mirhakkak et al. in the publication entitled "Dynamic Quality-of-Service for Mobile Ad Hoc Networks," MITRE Corp., 2000. Zhu discusses establishing bandwidth guaranteed QoS routes in small networks whose topologies change at a low to medium rate. Mirhakkak et al. are concerned with resource reservation requests which specify a range of QoS values while the network makes a commitment to provide service within this range.

**[0011]** Since MANETs are still in the initial stages of development, most attempts to implement QoS functionality in MANETs thus far have focused primarily on using QoS parameters to establish routes, as is the case with some

of the above-noted prior art approaches. Yet, as MANETs continue to increase in size and complexity, further QoS functionality may be needed along with ways to level loads in the network to reduce delays and optimize energy usage.

#### **Summary of the Invention**

**[0012]** In view of the foregoing background, it is therefore an object of the present invention to provide a method, mobile node and mobile ad hoc network that supports end-to-end delay reduction, QoS and energy usage leveling.

**[0013]** This and other objects, features, and advantages in accordance with the present invention are provided by a method for routing message data from a source node to a destination node in a mobile ad hoc network (MANET). The MANET has a plurality of intermediate mobile nodes between the source node and the destination node, and a plurality of wireless communication links connecting the nodes together. The method includes prioritizing the message data according to a type-of-service (ToS), discovering routes from the source node to the destination node, and ranking the discovered routes according to quality of service (QoS). Message data is distributed to the destination node on the discovered routes based upon the ToS of the message data and the QoS of the discovered routes, including distributing message data having a same ToS on multiple discovered routes, and distributing message data having higher priority ToS, e.g. delay sensitive message data and/or large volume message data, on higher ranked discovered routes.

**[0014]** Ranking the discovered routes according to QoS may include determining whether intermediate mobile nodes on discovered routes between the source node and the destination node are service sensitive nodes, such as power critical nodes and/or traffic bottleneck nodes. Also, ranking the discovered routes according to QoS may include determining an end-to-end delay of each discovered route, and may include measuring at least one of link delay, node capacity, node available capacity, and link reliability. Route entries may be stored in a route cache, each route entry corresponding to one of the discovered routes.

**[0015]** Objects, features, and advantages in accordance with the present invention are also provided by a mobile node for use in a mobile ad hoc network. The mobile node includes a communications device to wirelessly communicate with other nodes of the plurality of nodes via the wireless communication links, and a controller to route communications via the communications device. The communications include message data prioritized according to a type-of-service (ToS). The controller includes a route discovery unit to discover routing to a destination node, a route ranking unit to rank discovered routes according to quality of service (QoS), and a message data distribution unit to distribute the message data to the destination node along a plurality of the discovered routes based upon the ToS of the message data and the QoS of the discovered routes so that message data having a same ToS is distributed on multiple discovered routes, and message data having higher priority ToS is distributed on higher ranked discovered routes. Message

data having higher priority ToS includes delay sensitive message data and/or large volume message data.

**[0016]** The route ranking unit preferably determines whether intermediate mobile nodes on discovered routes between the source node and the destination node are service sensitive nodes, such as power critical nodes and/or traffic bottleneck nodes. The route ranking unit may determine an end-to-end delay of each discovered route, and may measure link delay, node capacity, node available capacity, and/or link reliability. The route ranking unit preferably includes a route cache to store route entries corresponding to one of the discovered routes.

#### **Brief Description of the Drawings**

**[0017]** FIGs. 1-3 are schematic diagrams of a mobile ad hoc network in accordance with the present invention.

**[0018]** FIG. 4 is a flowchart illustrating the method steps for a mobile ad hoc network in accordance with the present invention.

**[0019]** FIG. 5 is a schematic diagram illustrating a router of a node in accordance with the network of the present invention.

**[0020]** FIG. 6 is a schematic diagram illustrating the details of the controller of the router in FIG. 5.

#### **Detailed Description of the Preferred Embodiments**

**[0021]** The present invention will now be described more fully hereinafter with reference to the accompanying drawings, in which preferred embodiments of the invention are shown. This invention may, however, be embodied in many different forms and should not be construed as

limited to the embodiments set forth herein. Rather, these embodiments are provided so that this disclosure will be thorough and complete, and will fully convey the scope of the invention to those skilled in the art. Like numbers refer to like elements throughout, and prime notation is used to indicate similar elements in alternative embodiments.

**[0022]** As will be appreciated by those skilled in the art, portions of the present invention may be embodied as a method, data processing system, or computer program product. Accordingly, these portions of the present invention may take the form of an entirely hardware embodiment, an entirely software embodiment, or an embodiment combining software and hardware aspects. Furthermore, portions of the present invention may be a computer program product on a computer-usable storage medium having computer readable program code on the medium. Any suitable computer readable medium may be utilized including, but not limited to, static and dynamic storage devices, hard disks, optical storage devices, and magnetic storage devices.

**[0023]** The present invention is described below with reference to flowchart illustrations of methods, systems, and computer program products according to an embodiment of the invention. It will be understood that blocks of the illustrations, and combinations of blocks in the illustrations, can be implemented by computer program instructions. These computer program instructions may be provided to a processor of a general purpose computer, special purpose computer, or other programmable data processing apparatus to produce a machine, such that the instructions, which execute via the processor of the



computer or other programmable data processing apparatus, implement the functions specified in the block or blocks.

**[0024]** These computer program instructions may also be stored in a computer-readable memory that can direct a computer or other programmable data processing apparatus to function in a particular manner, such that the instructions stored in the computer-readable memory result in an article of manufacture including instructions which implement the function specified in the flowchart block or blocks. The computer program instructions may also be loaded onto a computer or other programmable data processing apparatus to cause a series of operational steps to be performed on the computer or other programmable apparatus to produce a computer or other programmable apparatus implemented process such that the instructions which execute on the computer or other programmable apparatus provide steps for implementing the functions specified in the flowchart block or blocks.

**[0025]** Referring initially to FIG. 1-4, a method for routing message data from a source node **S** to a destination node **D** in a mobile ad hoc network **20** will now be described. The network **20** includes a plurality of mobile nodes **30** including the source node **S** and the destination node **D** with intermediate nodes **IN** therebetween. The nodes **30**, such as laptop computers, personal digital assistants (PDAs) or mobile phones, are connected by wireless communication links **32** as would be appreciated by the skilled artisan. The method begins (block **100**) and includes prioritizing the message data according to a type-of-service (ToS), as indicated at block **102** in FIG. 4. At block **104**, routes are discovered

from the source node **S** to the destination node **D**, and then the discovered routes are ranked (block **106**) according to quality of service (QoS).

**[0026]** Message data is distributed (block **108**) to the destination node on the discovered routes based upon the ToS of the message data and the QoS of the discovered routes, including distributing message data having a same ToS on multiple discovered routes, and distributing message data having higher priority ToS, e.g. delay sensitive message data and/or large volume message data, on higher ranked discovered routes, as will be described in further detail below. As used herein, "message data" is intended to include any data that may be sent between nodes in the mobile ad-hoc network, including (but not limited to) video data, audio data, alphanumeric data, etc., as would be appreciated by the skilled artisan.

**[0027]** Ranking the discovered routes according to QoS may include determining whether intermediate mobile nodes **IN** on discovered routes between the source node **S** and the destination node **D** are service sensitive nodes (block **110**), such as power critical nodes and/or traffic bottleneck nodes. Also, ranking the discovered routes according to QoS may include determining an end-to-end delay of each discovered route (block **112**), and may include measuring at least one of link delay, node capacity, node available capacity, and link reliability. Route entries may be stored in a route cache, each route entry corresponding to one of the discovered routes. The method of the present invention may still follow the conventional procedures established by the protocol for best effort service.

**[0028]** Figs. 1-3 illustrates three scenarios involving load leveling. The scenarios include an individual concept, two possibly complimentary concepts and two competing concepts. The first scenario, Fig. 1, illustrates performance enhancement without regard to QoS. In this case, the source node **S** generates traffic or message data **A**, **B** with the same Type of Service (ToS) and destination node **D**. There are two discovered routes with similar or different QoS and the message data **A**, **B** is split up to reduce the delay for both data flows. In other words, message data **A** and message data **B** are sent over separate routes to the destination node **D**.

**[0029]** In the second scenario, Fig. 2, again the source node **S** generates traffic or message data **A**, **B** with the same destination node **D** but this time having different ToS. Again, there are two routes but in this example, one route has a lower QoS than the other. For example, here, the second route includes more hops and is longer. The higher priority message data **A** is sent along the shorter route and the lower priority message data **B** is sent along the longer route. This is meant to favor the higher ToS, but it could potentially improve performance for both flows.

**[0030]** The third scenario, Fig. 3, illustrates the use of load leveling for energy usage leveling in the network **20**. Here the source node **S** has two discovered routes to choose from but chooses not to use the shorter route because the shorter route has a service sensitive or power limited node **X** that is critical to one or more other routes, and there is another route without this service/power issue. These scenarios are not exhaustive but are meant to demonstrate that these three goals can

be independent, cooperative or competitive and that load leveling should be applied in a manner which looks at all three in an integrated way.

**[0031]** The described method can be applied to any type of On-Demand or Reactive routing protocol, such as Dynamic Source Routing (DSR) or Ad-Hoc On-Demand Distance Vector (AODV) routing, or to any hybrid proactive/reactive protocol, such as Zone Routing Protocol (ZRP), as would be appreciated by the skilled artisan. The described procedures are easily applied to the DSR protocol. The conventional DSR message types RREQ, RREP, RRER are defined as optional packet types, and can be used as defined for the conventional operation of the protocol to support "best effort" traffic in a backwards compatibility mode. New optional packet types may be defined to support the measurement of certain link metrics. Definition of the required header fields for these types is straightforward based on the functions defined above.

**[0032]** As discussed, the present invention provides the discovery and use of multiple routes. The routing cache/table is tailored to utilize up to  $n$  routes simultaneously per destination. Regarding the use of metrics: Improved metrics can include a variety of measures such as link delay, link/node capacity, link/node available capacity, link reliability. More important metrics may depend on the traffic class such as: Delay sensitive traffic may require routes to be rank-ordered based on the delay metric; Large volume traffic may require routes to be rank-ordered based on the capacity metric.

**[0033]** Thus, as described, the present invention supports conventional ad hoc routing protocols such as DSR and AODV but significantly improves performance. Multiple routes can be used for load balancing, and the use of multiple routes can provide more timely backup routes when path failures occur.

**[0034]** Referring now additionally to Figs. 5 and 6, a system aspect of the invention will be described. A mobile ad hoc network **20** includes a plurality of mobile nodes **30**, and a plurality of wireless communication links **32** connecting the plurality of mobile nodes together. Each mobile node includes a router **40** (FIG. 5) that has a communications device **42** to wirelessly and uni-directionally or bi-directionally communicate with other nodes via the wireless communication links **32**, and a controller **44** to route communications via the communications device **42**.

**[0035]** Also, a memory **46** may be included as part of the controller **44** or in connection with the controller. The communications include message data prioritized according to a type-of-service (ToS).

**[0036]** As shown in FIG. 6, the controller **44** includes a route discovery unit **50** to transmit route requests to other nodes to discover routing to a destination node. A route ranking unit **54** ranks discovered routes according to QoS, and a message data distribution unit **58** distributes the message data to the destination node along a plurality of the discovered routes based upon the ToS of the message data and the QoS of the discovered routes so that message data having a same ToS is distributed on multiple discovered routes, and message data having higher priority ToS is distributed on higher

ranked discovered routes. Message data having higher priority ToS includes delay sensitive message data and/or large volume message data. Again, the Qos ranking preferably includes a measurement of link delay, link capacity, link available capacity, and/or link reliability.

**[0037]** The message data distribution unit **58** distributes different message data along each of the plurality of discovered routes. The route ranking unit **54** comprises a route cache **56** for storing route entries each corresponding to one of the discovered routes. The ranking unit **54** may repeat route entries in the route cache for different classes of message data, such as delay sensitive message data and large volume message data.

**[0038]** The route ranking unit **54** preferably determines whether intermediate mobile nodes on discovered routes between the source node and the destination node are service sensitive nodes, such as power critical nodes and/or traffic bottleneck nodes. The route ranking unit **54** may determine an end-to-end delay of each discovered route, and may measure link delay, node capacity, node available capacity, and/or link reliability. The route ranking unit **54** includes a route cache **56** to store route entries corresponding to one of the discovered routes.

**[0039]** Many modifications and other embodiments of the invention will come to the mind of one skilled in the art having the benefit of the teachings presented in the foregoing descriptions and the associated drawings. Therefore, it is understood that the invention is not to be limited to the specific embodiments disclosed, and

that modifications and embodiments are intended to be included within the scope of the appended claims.